

What is claimed is:

1. A load balancer connected to a network connecting a plurality of clients requesting services and a plurality of servers executing operations based on said requests from said clients and replying with processing results comprising:

means for examining header information in request data from said client;

means for estimating, based on said header information and contents of said request data, processing load resulting from execution by said servers;

means for storing totals of said load estimates over a fixed past period for each of said servers;

means for dynamically selecting a server to which said request data is to be sent based on estimates of processing load on said servers resulting from current request data and total load for said servers; and

means for forwarding said request data to said servers.

2. A load balancer as described in claim 1 further comprising:

means for identifying a requested service type from said header of said request data; and

means for estimating processing load on said servers based on said service type.

3. A load balancer as described in claim 1 further comprising:

means for calculating requested data size based on said request data header and information about content data in said servers; and

means for estimating processing load on said servers based on said request data size.

4. A load balancer as described in claim 1 further comprising:

means for identifying program types to be executed by said servers based on said

request data header; and

means for estimating processing load on said servers based on execution of said programs.

5. A server load estimation method using an information processing device connected to a server and a client sending a service request packet to said server comprising the following steps:

requesting access to all services and all content data that can be provided by said server;

measuring processing load on said server associated with said request; and

generating data used to estimate, using said measurement results, server load resulting from request data from said client based on a header of said request data.

6. A load estimation method as described in claim 5 wherein, in said step for measuring processing load on said server, server processing load is estimated by measuring response time between when said client sends said service request packet and when a service response packet is received.

7. A load estimation method as described in claim 5 wherein, in said step for measuring processing load on said server, server processing load is estimated by measuring CPU load when said server receives said service request packet and executes an operation based on said request.

8. A computer-readable storage medium storing a program for implementing a method for estimating server load using an information processing device connected to a server and a client for sending a service request packet to said server, said method including the following steps:

requesting access to all services and all content data that can be provided by said server;

measuring processing load on said server associated with said request; and
generating data used to estimate, using said measurement results, server load
resulting from request data from said client based on a header of said request data.

9. A load balancing method using a processing device connected to a network connecting a plurality of clients requesting services and a plurality of servers executing operations based on said requests from said clients and replying with results from said operations, said method comprising the following steps:

examining header information in request data from said clients;

estimating, based on said header information and contents of said request data,
processing load resulting from execution by said servers;

storing totals of said load estimates over a fixed past period for each of said
servers;

selecting dynamically a server to which said request data is to be sent based on
estimates of processing load on said servers resulting from current request data and total
load for said servers; and

forwarding said request data to said servers.

10. A computer-readable storage medium storing a program for implementing a method for
estimating server load using an information processing device connected to a plurality of
clients requesting services and a plurality of servers executing operations based on
requests from said clients and replying with results from said operations, said method
including the following steps:

examining header information in request data from said clients;

estimating, based on said header information and contents of said request data,
processing load resulting from execution by said servers;

storing totals of said load estimates over a fixed past period for each of said
servers;

selecting dynamically a server to which said request data is to be sent based on

estimates of processing load on said servers resulting from current request data and total load for said servers; and

forwarding said request data to said servers.